DOCUMENT RESUME

ED 470 101 IR 021 539

AUTHOR	McKlin, Tom; Harmon, S.W.; Evans, William; Jones, M.G.
TITLE	Cognitive Presence in Web-Based Learning: A Content Analysis of Students' Online Discussions.
PUB DATE	2001-11-00
NOTE	8p.; In: Annual Proceedings of Selected Research and Development [and] Practice Papers Presented at the National Convention of the Association for Educational Communications and Technology (24th, Atlanta, GA, November 8-12, 2001). Volumes 1-2; see IR 021 504.
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Cognitive Processes; *Computer Mediated Communication; Computer Networks; *Computer Software Evaluation; Computer Uses in Education; *Content Analysis; *Discussion (Teaching Technique); Distance Education; Group Discussion; Online Systems; *Web Based Instruction; World Wide Web
IDENTIFIERS	*Neural Networks

ABSTRACT

This first phase of a content analysis of online, asynchronous, educational discussions is designed to generate a method for automatically categorizing messages into cognitive categories using neural network software. This phase of research answers two questions regarding the method of automatically analyzing discussion messages: Can a neural network reliably categorize messages under optimum circumstances, and how can the method be improved to generate great reliability? To determine whether neural network software can reliably categorize messages, two trials were conducted. The first, "best fit" trial, a proof of concept trial comprised only of messages which best fit the categorization model, generated strong reliability figures, and the second, systematic sample, a sample much more indicative of the messages generated in an online educational discussion, produced formative reliability figures from which the method of analysis may be optimized. This analysis also provides a distribution based on cognitive presence categories and subcategories of one semester of graduate online educational messages. (Contains 14 references.) (Author/AEF)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 470 101

Cognitive Presence in Web-Based Learning: A Content Analysis of Students' Online Discussions

Tom McKlin, S.W. Harmon, William Evans, and M.G. Jones

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

P. Harris

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

BEST COPY AVAILABLE

IR 021 539

Cognitive Presence in Web-Based Learning: A Content Analysis of Students' Online Discussions

Tom McKlin
Georgia State University/Georgia Institute of Technology

S.W. Harmon
William Evans
Georgia State University

M.G. Jones
Winthrop University

Abstract

This first phase of a content analysis of online, asynchronous, educational discussions is designed to generate a method for automatically categorizing messages into cognitive categories using neural network software. This phase of research answers two questions regarding the method of automatically analyzing discussion messages: Can a neural network reliably categorize messages under optimum circumstances, and how can the method be improved to generate greater reliability? To determine whether neural network software can reliably categorize messages, two trials were conducted. The first, "best fit" trial, a proof of concept trial comprised only of messages which best fit the categorization model, generated strong reliability figures (CR = 0.84; k = 0.76), and the second, systematic sample, a sample much more indicative of the messages generated in an online educational discussion, produced formative reliability figures (CR = 0.68; k = 0.31) from which the method of analysis may be optimized. This analysis also provides a distribution based on cognitive presence categories and subcategories of one semester of graduate online educational messages.

Many universities and K-12 educational settings have adopted online, web-based instruction as a tool for delivering instruction. According to Green (2000, para 7), "Today, 75 percent of two- and four-year colleges offer some form of online education. By next year, that number will reach 90 percent." Hamm (2000, para 8) makes a slightly more conservative claim by quoting a study performed by the Chronicle of Higher Education: "60% of American colleges and universities offer online-learning programs, and 8% more plan on doing so in the next year." He also notes that the e-learning market is expected to grow from $1.2 billion in 2000 to $7 billion in 2003. Certainly, online delivery of instruction is growing as are fora whereby students engage each other. WebCT, one of the more popular suite of tools to support web-based collaborative learning boasts 1600 new installations in the past 18 months and nearly 11 million student accounts (Goldberg, para 3). Although there is no clear data on the number of students participating in online courses in which every transaction is electronic, there appears to be a migration away from courses delivered solely face-to-face to those either supplemented with or completely reliant on online discussion. This migration toward electronic classrooms means that the discourse from these learning environments is very easily captured providing an opportunity for researchers to study the process of learning in a way that has never been available before. Never before have we had access to electronic texts containing virtually every exchange made by every student for an entire term. Concurrently, our ability to use computers to process text and reveal underlying themes has steadily grown (Rife, Lacy, & Fico, 1998). The convergence of these two realities brings us to our current state in which we have numerous texts available, a growing set of analysis tools, but very little research to explain the phenomena that take place in the course of learning. Kuehn (1994, p. 172) also highlights this dilemma, "few researchers have adopted current communication theory to investigate computer impact or effects in instructional settings...."

Despite the availability of electronic discussion list texts, few analyses of the content generated by students have been conducted. A content analysis type of inquiry allows us to describe how students engage and generate material within an online setting thereby providing potential answers to questions such as: Does a chatroom conversation produce different cognitive results than either a teacher-led asynchronous discussion or a student-led asynchronous discussion? Henri (1992) makes apparent the role content analysis has to play in an instructor's ability to guide learning:

BEST COPY AVAILABLE

> Content analysis, when conducted with an aim to understanding the learning process, provides information on the participants as learners, and on their ways of dealing with a given topic. Thus informed, the educator is in a position to fulfill his main role, which is to offer immediate support to the individual and the collective learning process. (p. 118)

Over all, this study outlines the initial phase of the construction and use of a neural network to perform a content analysis of a large body of student messages for cognitive presence, one portion of Garrison, Anderson, and Archer's (2000) model to understand online learning environments. This type of tool may ultimately be used to gauge, guide, direct, and manipulate the learning environment. Despite Howell-Richardson & Mellar's (1996) research indicating that modifications to the structure of an online course produce significantly different communication outcomes, instructors currently have little ability to gain a bird's-eye view of the overall learning taking place, much less an ability to respond to that learning, assess it, or intervene. This research seeks to answer two questions. First, can neural networks be used to analyze and describe the cognitive landscape of online educational discussions? Second, at this phase, how is cognitive presence displayed in an online course?

Theoretical Background

Cognitive Presence

Garrison, Anderson, and Archer (2000, 2001) have developed a community of inquiry model, based on Dewey's (1933) practical inquiry model, which splits community-based learning into three overlapping areas: social presence, cognitive presence, and teacher presence. They operationalize cognitive presence by splitting it into four phases: triggering event, exploration, integration, and resolution, and use the following descriptors respectively for each phase: evocative, inquisitive, tentative, and committed. Specifically, cognitive presence is defined as "the extent to which learners are able to construct and confirm meaning through sustained reflection and discourse in a critical community of inquiry" (p. 11). Garrison et al. employ their cognitive presence model to analyze an online discussion group. Their unit of analysis is the entire message mainly because messages are easiest to identify and occur naturally in discussion environments. Because a message may contain indicators for multiple phases, they have developed two heuristics for deciding which messages fall into which categories: code down and code up. They used human coders to classify messages, and this yields a reliability figure (k=0.74) which Riffe, Lacy, and Fico (1998) accept only for research that is breaking new ground, a category under which this research clearly fits. Also, they found that the greatest coding discrepancies occurred between coding for exploration and integration. They admit low occurrences of resolution and believe higher instances of resolution will be found "where applied knowledge is valued—particularly adult, continuing, and higher education" (p. 16)

Can Neural Networks Analyze Messages?

The use of neural networks in educational settings is rare and there are no accounts outlining the use of a neural network to analyze text messages of an online discussion group. Garson (1998) provides a number of reasons why social scientists have not adopted the use of neural networks in their research. First, neural network software has been available to the social scientist only since the early 1990's. Second, it is not clear how neural networks arrive at their conclusion; unlike an expert system, neural networks provide no audit trail outlining their reasoning. Also, neural network techniques are complex and leave researchers unsure whether their analysis is truly optimal; slight modifications to a number of parameters may yield a more optimal analysis. Nonetheless, neural networks are good at making predictions (e.g. stock market forecasting) and at classification. Garson cites 34 research studies using neural networks in economics and business, 9 in sociology, 7 in political science, and 45 in psychology (Garson, 1998, pp. 8-22).

Given a high enough reliability value, neural networks have the ability to classify large quantities of data which, for the present study, means that researchers do not have to sample a subset of all online messages from a course. Instead, the neural network classifies each message thereby eliminating sampling error. In comparison with statistical methods of analysis, Garson (1998) mentions:

> [N]eural models may outperform traditional statistical procedures where problems lack discernible structure, data are incomplete, and many competing inputs and constraints related in complex, nonlinear ways prevent formulation of structural equations, provided the researcher can accept the approximate solutions generated by neural models (p.1)

Clearly, student messages are filled with competing inputs related in a complex, nonlinear fashion. Further, traditional textual analysis of this type would require the use of multiple human coders classifying each message against a set of classification criteria, a resource-intensive technique which also generates approximate solutions.

Method

The method involves four steps starting with a text-based transcript of an online discussion and ending with the calculation of reliability statistics.

BEST COPY AVAILABLE

Database Creation

First, one semester's worth of asynchronous, online discussion messages were converted from a single text file containing all messages for one semester into a database such that each record represented one message and contained the message body, author, date, etc. This task was accomplished using SQL Server and a series of SQL statements to populate the database. These generic tools were used to streamline the process of making them publicly available over the World Wide Web.

Word Count Tool

Second, a tool was constructed to page through each message body and perform word counts in both self-defined and General Inquirer categories (see Danielson & Lasorsa, 1997; http://www.wjh.harvard.edu/~inquirer). The categorical word count procedure results in a database table with categories as columns, individual messages as records, and cell values representing the count of terms from each cognitive presence category. Self-defined categories allow researchers to define specific indicators for each category. For example, items falling into the cognitive presence phase "integration" often refer to previous messages or draw from a course participant's prior knowledge; therefore, typical "integration" messages incorporate terms and phrases such as "thanks," "that reminds me of," "compared to," and "I agree." The researcher may create one or a number of categories that serve as indicators that a message should be categorized as an integration message. This tool not only allows for the creation of new, user-defined input categories but also incorporates existing input categories from the dictionary of terms found in the General Inquirer. This dictionary is comprised of 11,788 words in 182 categories. Each message was analyzed against each self-defined and General Inquirer category of terms and a simple word count was taken to determine the weight of each category of terms in each message. For example, the General Inquirer category "positiv" contains the words "up, abide, and yes" meaning that the following sentence will receive a "positiv" score of two: "Yes, I had to look up to see the icon." Further, the "positiv" score of 2 is normalized so the neural network can accurately compare scores across messages. Normalization is performed by dividing the number of times the terms in a single category appear in a message by the total number of words in the message (2/10 = 0.2).

Neural Network Training

Third, a feedforward, backpropagation, neural network was trained to classify each message as falling into one of five categories (triggering event, exploration, integration, resolution, or noncognitive). This was done by human-classifying a group of messages to be used as the training set, training the neural network on that set of messages, and then classifying a second set of messages for reliability purposes.

Reliability Measures

Fourth, reliability measures were taken comparing human-coded messages with those classified by the neural network. Huck (2000) recommends the use of multiple reliability measures for a single study (p. 98). For this reason and because this study replicates a similar study by Garrison, Archer, and Anderson (2001), two reliability measures were employed: Holsti's (1969) coefficient of reliability (CR) which measures the agreement between two coders divided by the total number of messages analyzed and Cohen's kappa which corrects for chance agreement among coders. The difference between the Garrison et al. study and this one is that Garrison et al. performed a human – human comparison whereas this study performed a neural network – human comparison.

Results

To determine whether the neural network analysis produces results comparable to human-coded content analysis, benchmarks from a human-coded content analysis by Garrison, Anderson, and Archer (2001) were compared to results from this neural network analysis. Garrison et al. went through three phases of training human coders to reliably categorize messages and used both Holsti's (1969) coefficient of reliability (CR) and Cohen's (1969) kappa (k) to measure inter-rater reliability. Garrison et al. generated the following reliability figures:

Reliability Measure	Trial 1	Trial 2	Trial 3
CR	0.45	0.65	0.84
Kappa	0.35	0.49	0.74

Table 1. Reliability measures for Garrison, Anderson, & Archer's (2001) content analysis of an online discussion.

Best Fit Sample

The analysis using the neural network to classify messages is a multi-phase process of which this paper presents the first phase. This phase seeks to answer whether it is possible at all for a neural network to classify messages. In this phase, the messages best representing each category were coded and used to train and test the neural network model. This "best fit" trial yielded the following reliability figures: CR = 0.84 and k = 0.76. This test set (n=26) of optimal messages generates the following matrix after being run through the trained "best fit" model.

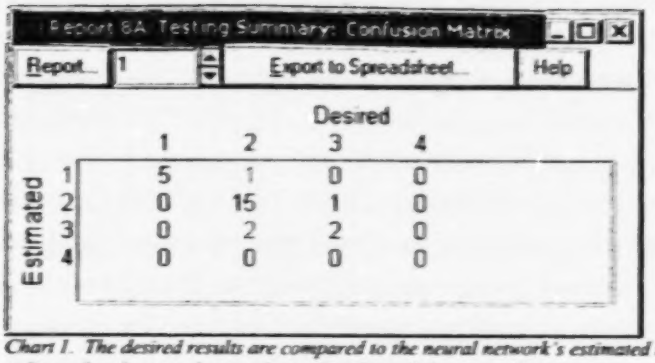

Chart 1. The desired results are compared to the neural network's estimated results, and the numbers appearing diagonally indicate that the neural network matched the coded test set of messages.

In this set, 1 indicates a triggering event, 2 is an exploratory message, 3 is an integration message, and 4 is a resolution message. This trial indicates that a neural network can reliably discern the first three categories.

Systematic Sample

The purpose of the "best fit" trial is to determine whether a neural network can be used to categorize text messages at all; the second trial uses a systematic sample of messages in which both the training set (n=100) and the test set (n=100) are a systematic sample of every 20 messages. There are 1,997 messages in all; therefore, this sample represents a cross-section of messages occurring throughout the term. Further, this sampling technique introduces noise into the analysis; to accommodate for this, a fifth (noncognitive) category was used. This category represents non-cognitive messages (e.g. greetings and short agreement messages), course management messages (e.g. "When will the textbook be available," or "when is the next chat?"), and technical support messages (e.g. " I can't get into the chat room," and "Why are my messages not showing up on the discussion list?"). A neural network trained against 100 messages using all five categories yielded a CR value of 0.68 and a kappa value of 0.31 generating the following message category results:

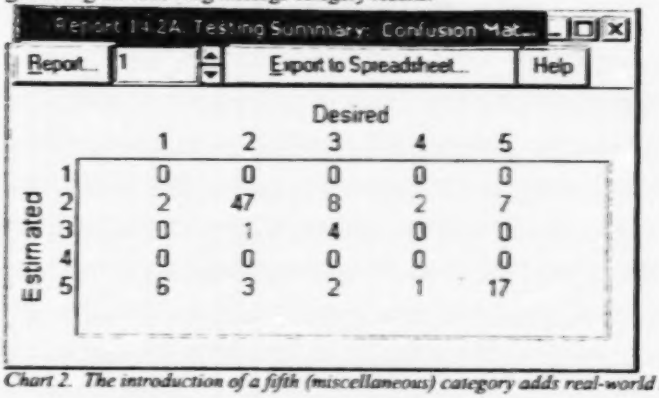

Chart 2. The introduction of a fifth (miscellaneous) category adds real-world noise to the neural network analysis.

Finally, the systematic sampling of 20% of the messages from the term provides insight into the cognitive effort displayed by the course participants. The first chart displays the percentage of messages by broad cognitive category type, and the second displays messages by the subcategories which make up each category type.

Cognitive Presence by Category	Percentage
Not Cognitive	48%
Triggering Event	3%
Exploration	39%
Integration	9%
Resolution	1%

Table 2. Messages by cognitive presence category.

BEST COPY AVAILABLE

Cognitive Presence By Subcategory	Percentage
Not Cognitive	
Unrelated	20%
Course Management	12%
Technical Support	12%
External Reference	4%
Total Not Cognitive	48%
Triggering Event	
Recognizes Problem	1.5%
Puzzlement	1.5%
Total Triggering Event	3%
Exploration	
Personal Narrative	6%
Information Exchange	12%
Brainstorming	5%
Divergence Among	4%
Leap to Conclusion	5%
Suggestion	6%
Divergence Within	1%
Total Exploration	39%
Integration	
Creating Solutions	0.5%
Synthesis	1.5%
Convergence Within	2%
Convergence Among	5%
Total Integration	9%
Resolution	1%

Table 3. Messages by cognitive presence subcategory.

Discussion

Findings

The first trial indicates that in the absence of noise, a neural network can be used to categorize messages into the cognitive categories outlined by Garrison, Anderson, and Archer (2001) based on linguistic cues. In the trial which introduced the noise which naturally occurs in discussion lists, we see that the model overgeneralizes on categories two (exploration) and five (miscellaneous), that it undergeneralizes on integration messages, and that it does not discern triggering events and resolution messages from the others. These findings provide critical, formative information which can be used to optimize and therefore improve the model. Methods for improving the model's ability to correctly categorize are outlined below.

Optimization

Just as Garrison et al.'s coders optimized their coding algorithm between times they coded, the neural network method of analysis may also be optimized. The above reliability reflects an initial brute-force analysis of each message and takes as input weights generated by analyzing each message against a category of terms in the General Inquirer dictionary. The following steps may be taken to improve the model:

Word sense disambiguation: This simply means that individual words are classified according to their parts of speech. For example, the word "test" may be used as either a verb or a noun, and a word sense disambiguation routine will clearly separate those instances of "test" that are nouns and those that are verbs. This should dramatically reduce the amount of noise in the database.

Increased training set: The next phase of this research is the analysis of six eCore courses, online post-secondary, core curriculum courses offered by the University System of Georgia. In this phase of research, six instructors will analyze 200 messages each thereby generating a training set of 1100 messages. In comparison, the current research used 100 messages as its training set and 100 messages as a test set. Building a model from 1100 coded messages should improve the generalizability of each category and therefore the model's ability to correctly classify messages.

Message hierarchy metainformation: In the current model, the only hierarchy information fed into the neural network is whether each message is a reply to another message or not. Garrison et al.'s model indicates that messages are partially classified not only based on their textual content but on their place within a given thread hierarchy. If a message is the first in it's thread

BEST COPY AVAILABLE

hierarchy, it is most likely a triggering event. If it is near the beginning of a thread and is a response to another message, it is most likely either an exploratory or integrative message.

Improved categories: Create subsets of each category that are very specific, and ensure that each message fits cleanly into a category.

Cognitive Presence Distribution

The distribution of messages into cognitive presence categories is similar to that found by Garrison, Anderson, and Archer (2001) in that a majority of messages fell into the exploration category with fewer integration messages, only a few triggering events, and practically no resolution messages. The discussion topics and goals of the course define the distribution we found. The goal of the course is to give each student experiential knowledge of web-based learning. It is up to the instructor to define whether resolution can practically be achieved in the course; resolution is usually reserved for more practical tasks in which students state that they have resolved an issue which means they have applied knowledge in a real-world setting and have found that the real-world outcome affirms knowledge gained from the course. Although students were creating their own web-based learning modules, these modules were not intended to be the product of a learned body of knowledge; rather these modules were intended to be tasks from which questions emerge. Given this course structure, it makes sense that resolution is rare and exploration dominates. Interestingly, the number of triggering events is fairly low which may also be attributed to the course structure; students were not given a formal triggering event or question by the instructor each week; instead, the instructors allowed the students' exploration to define the direction of the course. In this case, triggering events were more likely to be found embedded within exploratory messages. Tracing triggering events may be assisted by the creation an overall diagram of the course structure allowing us to see not triggering events as defined by the linguistic cues within the message but rather as defined by the messages emanating from these triggering events. That is, if we find that one message spawns a critical debate, then we may in retrospect define that message as a triggering event. This information can be displayed graphically for use by those coding the training set of messages and numerically for use by the neural network.

The Next Phases of this Research

It is expected that a well-trained neural network will perform just as reliably as a set of human coders at classifying messages into cognitive presence categories. This method of analysis will then provide a broad overview of the cognitive effort displayed by students throughout the semester and allows for instructors to make adjustments to their approach in order to bring about desired displays of cognitive effort. Further, this rapid analysis method provides a tool instructors may use to conduct their own research on finely grained aspects of the cognitive dynamics of a course. This method may allow us to answer questions such as: Which displays of instructor involvement generate exploration and which generate integration? Are socially engaged students also cognitively engaged? How many course participants is optimal for higher order thinking? Which class participants encourage the integration of ideas?

References

Danielson, W. A. & Lasorsa, D. L. (1997). Perceptions of social change: 100 years of front-page content in The New York Times and The Los Angeles Times. In Text Analysis for the Social Sciences: Methods for Drawing Inferences from Texts and Transcripts, edited by Carl W. Roberts, 103-115. Mahwah, NJ: Lawrence Erlbaum.

Dewey, J. (1933). How we think: A restatement of the relation of reflective thinking to the educative process. Boston: D. C. Heath.

Garson, G. D. (1998). Neural Networks: An Introductory Guide for Social Scientists. Thousand Oaks: Sage.

Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: computer conferencing in higher education. The Internet and Higher Education 2(2-3), 87-105.

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. American Journal of Distance Education 15(1), 7-23.

Goldberg, M. (2000). The year educational technologies "grew up." WebCT. Retrieved February 10, 2001, from the World Wide Web: http://www.webct.com/service/ViewContent?contentID=2665742&communityID=863.

Green, J. (2000). The online education bubble. The American Prospect 11(22), 32-35.

Hamm, S. (2000). The wired campus. Business Week 37(11), 104-112.

Henri, F. (1992). Computer conferencing and content analysis. In A. R. Kaye (Ed.), Collaborative learning through computer conferencing: The Najaden papers (pp. 115-136). New York: Springer.

Holsti, O. (1969). Content analysis for the social sciences and humanities. Don Mills, ON: Addison-Wesley.

Howell-Richardson, C. and Mellar, H. (1996). A methodology for the analysis of patterns of participation within computer mediated communication courses. Instructional Science 24, 47-69.

Huck, S. (2000). Reading Statistics and Research. New York: Longman.

Kuehn, S. A. (1994). Computer-mediated communication in instructional settings: A research agenda. Communication Education 43(2), 171-184.

Riffe, D., Lacy, S., and Fico, F. G. (1998). Analyzing media messages: Using quantitative content analysis in research. Mawah, New Jersey: Lawrence Erlbaum.

BEST COPY AVAILABLE

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

[X] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

[ ] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").